094164

REPORT TO THE SUBCOMMITTEE
ON FISHERIES AND WILDLIFE
CONSERVATION AND THE
ENVIRONMENT, COMMITTEE ON
MERCHANT MARINE AND FISHERIES
HOUSE OF REPRESENTATIVES

Federal Environmental
Data Systems B-177222

BY THE COMPTROLLER GENERAL
OF THE UNITED STATES

094164   NOV. 22, 1974

ROUTING SLIP

TAN: 710785
DATE: 6/6/80

Index KY 6/18/80	1st Catalog KY 6/18/80
Abstract	Initial Q.C.
Final Q.C.	2nd Catalog KY 6/18/80

B-177222

The Honorable John D. Dingell, Chairman
Subcommittee on Fisheries and Wildlife
  Conservation and the Environment
Committee on Merchant Marine and Fisheries
House of Representatives

Dear Mr. Chairman:

In response to your March 8, 1973, request, this is our
report on our review of environmental data collection and
storage activities in the Federal Government.

We do not plan to distribute this report further unless
you agree or publicly announce its contents.

Sincerely yours,

Comptroller General
of the United States

ABBREVIATIONS

CSC	Civil Service Commission
EDS	Environmental Data Service
ENDEX	Environmental Data Index
EPA	Environmental Protection Agency
FDA	Food and Drug Administration
GAO	General Accounting Office
NAWDEX	National Water Data Exchange
NIH	National Institutes of Health
NOAA	National Oceanic and Atmospheric Administration
OMB	Office of Management and Budget
RIAL	Resource and Land Information
R&D	research and development
SEQUIP	Study of Environmental Quality Information Programs
USGS	United States Geological Survey

COMPTROLLER GENERAL'S REPORT
TO THE SUBCOMMITTEE ON FISHERIES
AND WILDLIFE CONSERVATION AND
THE ENVIRONMENT
COMMITTEE ON MERCHANT MARINE
AND FISHERIES
House of Representatives

FEDERAL ENVIRONMENTAL DATA SYSTEMS
B-177222

D I G E S T

WHY THE REVIEW WAS MADE

The Subcommittee Chairman ex-
pressed concern that the Federal
Government was not making maximum
use of environmental data collected
by diverse Federal organizations.

He also indicated that, there was no
directory of environmental informa-
tion and that there may be little
coordination among such organiza-
tions in collecting and exchanging
data.

In line with the Chairman's request,
GAO examined into the

--kinds of environmental data col-
   lected and scope of data collec-
   tion programs;

--methods of information storage;

--availability of information to
   private and public sectors,
   including other government agen-
   cies; and

--methods of coordinating data col-
   lection and transfer among agen-
   cies. (See p. 1.)

FINDINGS AND CONCLUSIONS

Agencies collecting and
storing environmental data

GAO distributed a questionnaire to
managers of Federal systems which
collect, store, and disseminate
environmental data. (See p. 1.)
It received responses from managers
of 320 separate data systems in 8
major departments and 10 independent
agencies collecting and storing
environmental data, many on a
nationwide basis.

Many of these systems were collect-
ing and storing data in the same
functional areas. For example, 120
systems were collecting and storing
air data. Most systems were storing
data on some type of computer media,
such as punched cards and magnetic
tapes. (See p. 2.)

Previous studies have found that
environmental data systems were
developed to satisfy needs perceived
by individual organizations without
overall coordination between
agencies or systems.

For example, a 1971 study by the
Committee to Study Environmental
Quality Information Programs for the

Tear Sheet. Upon removal, the report
cover date should be noted hereon.

i

President's Office of Science and Technology noted that dispersed, uncoordinated and often overlapping information programs resulted from the fact that their primary tasks are to support the agencies' missions. Deliberate interaction with other information programs is often outside the scope of a given program. (See p. 3.)

Data exchange between Federal agencies

Although the Federal Government has no central focal point for exchanging or collecting environmental data, many systems are sharing data with other Federal agencies. Of the 320 systems, 167 were providing environmental data and information to other agencies and 94 were receiving data from systems in other agencies. (See p. 4.)

Managers of 98 systems indicated that they were participating in various networks which had as their purpose either the exchange of data or the coordination of collection activities.

In talking to some respondents, GAO found only one system transferring data automatically between computers. Other methods included

--exchange of computer media, such as magnetic tapes;

--use of remote terminals; and

--exchange of hard copy or written material.

Some formal interagency committees promote data exchange. For example, the Federal Working Group on Pest Management under the Council on Environmental Quality is concerned with coordinating various pesticide data collection activities of nine agencies. (See p. 5.)

Improvement in data exchange is being made through establishment of referral centers. The Department of the Interior is designing a National Water Data Exchange which is to maintain a computerized master index of water data sources to facilitate exchange of water data. (See p. 6.)

Various agency officials told GAO that, in the absence of a formal organization for coordinating data collection and exchange, some coordination was attained through informal professional contacts between personnel in various agencies. (See p. 6.)

Problems in exchanging data

Systems managers identified the following problems in acquiring and using data from other systems.

--Data elements and codes used in one system were not compatible with those used by another system.

--Responses to requests for data were often delayed.

--Accuracy and reliability of data collected by other systems was difficult to assess.

--Equipment and languages used in one system were not compatible with those used in another system.

--Recoding or converting data was expensive. (See p. 7.)

Directory of environmental data systems

Both the 1971 report by the Committee to Study Environmental Quality Information Programs and a report on the 1972 National Environmental Information Symposium sponsored by the Environmental Protection Agency noted that a basic problem of data exchange was the lack of awareness of what data had been collected and by whom. It appears the problem still exists, because managers of 151 systems expressed the need for an overall directory of environmental data systems.

They indicated a directory could

--eliminate duplication in data collection,

--reduce search time and the number of referrals, and

--identify existing information sources and what data is available. (See p. 11.)

GAO found that some agencies had prepared directories of their data systems. Some directories of environmental data sources beyond a single agency had been published.

However, GAO is unaware of any Federal agency maintaining a complete directory of all environmental data systems. (See p. 13.)

Interior is planning to develop directories that will list certain environmental data sources in

--its Resource and Land Information (RALI) program--land use and

natural resources information sources--and

--its National Water Data Exchange. (See p. 13.)

The National Referral Center of the Library of Congress maintains numerous references to environmental data sources. However, the Center has not issued a directory of all its environmental data sources and has no plans to do so. (See p. 14.)

Creating a directory of environmental data systems is one possible solution to the lack of awareness of what data has been collected and by whom.

Network of environmental data systems

Bills have been introduced in the Congress to establish a central organization to coordinate collection and exchange of environmental data. One bill--House bill 56--was passed by both Houses of the 92d Congress but was vetoed by the President. He said it would establish a central computer system which could be unnecessarily costly, because it would lead to duplication of information or would produce results unrelated to real needs.

Another bill introduced, but not acted upon, in the 93d Congress--House bill 36--would provide specifically for a network of new and existing data processing facilities which, through a system of interconnections, would be in communication with a central facility.

The Subcommittee Chairman asked GAO to look into costs and problems of

establishing an environmental data
network. (See p. 15.)

In commenting on House bill 36, GAO
said the bill could fulfill the need
for a national environmental data
system to coordinate independent
systems of many Federal agencies
conducting research and development
programs. In another review, GAO
found a lack of coordination among
agencies collecting and disseminating
water pollution research and
development information and recom-
mended designation of a focal point to
coordinate such information. (See
p. 17.)

The Subcommittee Chairman asked GAO to
develop costs for an existing computer
network and to relate such costs to
the cost of establishing a network of
environmental data systems.

According to one large Government-
owned computer network, however, costs
of an existing network cannot be
readily related to costs of establish-
ing a new network.

Costs of a new network are dependent
upon specific purposes of the network,
design of the network, and the
hardware and software needed to ac-
complish the network's objectives.
(See p. 18.)

There are less costly methods of
exchanging computerized data--such as
exchange of magnetic tapes--than
direct interconnection of computers.

GAO also looked into the problems of
setting up such a network. The De-
partment of the Interior's Geological
Survey considered establishing a data
network to support land use and
resource planning at the Federal,
State, and local levels.

As developed in a contractor's
report, the concept for the
network--referred to as the RALI
program--is similar to the concept in
House bill 36.

The RALI network would have made
maximum use of existing information
sources by providing

--references to data maintained by
  other organizations and

--an information base for modeling
  and analysis from data gathered
  and maintained by other organiza-
  tions.

A national information facility would
have been established to supply
analyzed and summarized data to the
Federal establishment. (See p. 18.)

The contractor's report suggested
that development of the RALI program
be based on a survey of users' needs
and data sources which could be
incorporated into the RALI program.

The Director of the RALI program told
GAO that the contractor's approach
could not be implemented because it
was data oriented rather than problem
oriented.

He explained that most data was
useful for the specific purpose it
was collected and might have only
marginal utility beyond its original
use. Introducing data into an
information system without regard to
its utility can be costly in the
collection and maintenance of that
data.

He added that networks should be
designed around problems to be solved
and not around miscellaneous, unre-
lated, and limited-use data collected

by various data users. (See p. 19.)

MATTERS FOR CONSIDERATION BY THE SUBCOMMITTEE

GAO suggests that the Subcommittee, when considering House bill 36, specify that, before a nationwide network of interconnected environmental data systems is established, the environmental problems to be solved be defined and the analysis tools needed to assist in solving the problems be determined. Once problems and analysis tools are identified, data needs for the network and the best methods for storing and exchanging such data can be determined.

GAO also suggests that the central organization to be established, if the bill were enacted, be responsible for establishing and maintaining an environmental data directory to increase awareness of available data.

CHAPTER 1

INTRODUCTION

In a letter of March 8, 1973 (see app. I), the Chairman, Subcommittee on Fisheries and Wildlife Conservation and the Environment, House Committee on Merchant Marine and Fisheries, said that environmental data and information were being collected by a number of diverse Federal organizations and that this data was a national resource representing an investment of several million dollars. The Chairman expressed concern that (1) the Federal Government was not making maximum use of this valuable resource, (2) there was no systematic directory of the environmental information collected and stored by the Federal Government, and (3) there may be little coordination among agencies in the collection and exchange of data.

In line with the areas of interest expressed in the Chairman's letter, we examined into the

--kinds of environmental data collected and scope of the data collection programs,

--methods of information storage,

--availability of information to the private and public sectors, including other Government agencies, and

--methods of coordinating environmental data collection and transfer among Government agencies.

SCOPE OF REVIEW

We distributed a questionnaire on the above four areas to the managers of Federal systems involved in collecting, storing, and disseminating environmental information and data. We talked to a number of respondents to our questionnaire to discuss or clarify the information furnished. In addition, at the Chairman's request, we obtained information on (1) the costs involved in establishing a computer network to facilitate the exchange of environmental data and (2) the problems of establishing such a network.

CHAPTER 2

MANY FEDERAL AGENCIES COLLECT

AND STORE ENVIRONMENTAL DATA

Many Federal departments and independent agencies collect and store environmental data. In many instances, the information is collected in the same functional areas, such as air and water, and is stored primarily on some type of computer media, such as punched cards and magnetic tape.

The environmental data systems appear to have been developed independently of each other to meet the specific missions of the various organizations. Although the Federal Government has no central focal point for coordinating the establishment of environmental data systems or for exchanging or collecting environmental data, many managers of environmental data systems told us that they did exchange data with other Federal agencies and, in some cases, coordinated the exchange and/or collection of data through some type of network. These managers said, however, that a number of problems hindered the exchange of data. These problems included a lack of uniformity in the way data is coded in the various computer systems and a lack of awareness of what data has been collected and by whom.

AGENCIES COLLECTING AND STORING
ENVIRONMENTAL DATA

We identified 8 major departments and 10 independent agencies that collect and store environmental data and information. We received responses to our questionnaire from officials who manage 320 separate environmental data systems. (See app. II for a summary of these responses.) Following are the significant departments and agencies in terms of numbers of systems.

Department or agency	Number of systems
Environmental Protection Agency	87
Interior	61
Commerce	39
Agriculture	33
Total	220

We found that many of these systems in a number of different agencies were collecting and storing data in the same functional areas. For example, 120 systems in 14 different agencies were collecting and storing air data. Also 175 systems in 17 agencies were collecting water data and 105 systems in 15 agencies were collecting land use data. (See p. 25 of app. II.)

Many of the systems were collecting environmental data on a nationwide basis. Water data and air data are collected throughout the United States by 58 and 30 systems, respectively.

For 229 systems the data is stored totally or in part on some type of computer media, such as magnetic tape, punched cards, magnetic disc, and drum storage. Other methods of storage included reports, publications, and other hard-copy forms; microfilms; and charts.

Environmental data systems collect and store data primarily to support functions in a specific agency. Respondents indicated that the primary purposes of the Federal systems' data collection efforts were to support

--management and planning;

--research and development;

--surveillance and monitoring; and

--legal, legislative, or regulatory functions.

Previous studies have found that data systems generally have evolved on the basis of needs perceived by individual organizations without overall coordination between agencies or systems.

In a 1971 study made by the Committee to Study Environmental Quality Information Programs (SEQUIP) for the President's Office of Science and Technology, it was noted that dispersed, uncoordinated, and often overlapping programs

"* * * result of the fact that, in most cases, the primary task of most information centers is to support the mission of their parent agencies. Deliberate interaction with other information programs is often outside the scope of a given program, * * *."

3

The SEQUIP Committee noted further that the formation of the Environmental Protection Agency (EPA) and the National Oceanic and Atmospheric Administration (NOAA), Department of Commerce, should alleviate some of the overlap in data collection and storage activities and should improve communications between environmental data systems having similar functions but in different agencies. However, in a February 1974 report to EPA on an analysis of EPA's scientific and technical information activities to serve as a basis for developing an information network, the Battelle Memorial Institute noted that:

> "There are a large number of heterogenous information/data activities dispersed irregularly throughout EPA today very much as activities were at the formation of EPA in December 1970. Many of these activities are not guided by national or agency environmental goals except to the extent that while satisfying the needs of their immediate funder, they also contribute to the achievement of national goals. Such systems are very vulnerable to parochial initiation and termination."

The report further noted that:

> "The present EPA scientific and technical information network generally is a free unorganized network. Such a network permits individual systems to develop whenever and wherever they are needed. To the extent that interaction or cooperation between the diverse systems is required, the operators of each of the individual systems must establish and maintain such channels of interaction as they see fit."

The report suggests the establishment of an office within EPA "to plan, coordinate and encourage improved accessibility, handling and usage of environmental information and data within a coordinated network." This office would perform an information coordination function in EPA by providing referrals to sources of scientific and technical information.

DATA EXCHANGE BETWEEN FEDERAL AGENCIES

Federal agencies are sharing some environmental data. Of 320 systems, 167 were providing data and information to other departments or agencies and 94 were receiving environmental data from systems in other departments or agencies. For example, the Corps of Engineers has a cooperative agreement with NOAA's National Ocean Survey to provide water

data. After the Corps' computer analyzes the data, it is given to the National Ocean Survey for use in the development and publication of charts for marine navigation.

EPA receives data on water quality and other measurements collected by the U.S. Geological Survey (USGS). Measurements are received from USGS on magnetic tapes. These measurements are another source of water quality data for EPA which receives similar data from various Federal and State organizations.

Managers of 98 systems indicated that they were participating in various networks which had as their purpose either the exchange of data or the coordination of collection activities.

We talked to a number of respondents who appeared to be exchanging data in a formal computerized network. However, we found only one network transferring data automatically between computers of different systems. The National Weather Service, a member of the World Meteorological Organization, participated in a network which transferred weather data between countries.

We did find a number of other methods used to exchange data. These methods included (1) the exchange of data in a computer media, such as punched cards, paper tapes, magnetic tapes, and discs, (2) the transfer of data over communication lines by use of a remote terminal, and (3) hard-copy exchange, such as reports and other written material. We do not know how often or how much environmental data is exchanged using each method. In fiscal year 1971, the General Services Administration completed a survey of 14 civil departments and agencies to measure existing and potential needs for data exchange as a basis for designing nationwide communications networks and systems. The General Services Administration found that the majority of reports--52,200 of 54,000, or 97 percent--were exchanged in hard-copy form and the balance were sent in a computer media.

Although there is no central focal point for collecting and transferring environmental data, there are some formal interagency committees which consider, among other things, various problems of data exchange.

For example, the National Pesticide Monitoring Program is a network of data collection systems established to monitor changes in the levels of pesticide residues in various environmental components, such as soil, air, water, and estuaries. The program was established in 1964 by the

Federal Committee on Pest Control which was chartered by the
Secretaries of four Federal departments. This interagency
committee has been expanded to include eight departments and
one agency and is now called the Federal Working Group on
Pest Management. Under the Council on Environmental Qual-
ity, this Working Group and its various panels handle co-
ordination problems between the departments. The Monitoring
Panel advises the Working Group on interagency coordination
matters related to the monitoring systems.

Agency officials told us that, where formal organiza-
tions did not exist for coordinating data collection and
exchange between systems collecting environmental data,
many informal professional contacts did exist between per-
sonnel in agencies.

The SEQUIP study noted, however, that "interaction
among similar, or logically related, programs are rare
and are usually not due to deliberate network planning."
It was found:

> "* * * during the SEQUIP Workshop that often infor-
> mation program managers working in very similar
> subject areas, but in different mission agencies,
> do not know of each other's activities. Thus, the
> discussions of the radiation panel (which covered
> only ten information programs) showed that the
> program representatives present at the session were
> not aware of the existence of one-third of the other
> programs * * *."

Improvements in the area of data exchange are being made
through the establishment of a referral center or a central
point to facilitate the location and dissemination of data.
The Department of Interior is designing a National Water Data
Exchange (NAWDEX) to improve the use of collected water
data. The NAWDEX concept provides for a "Systems Central"
which will compile and maintain a computerized master index
of water data and water data sources. NAWDEX's main func-
tion is to facilitate the transfer of water data from those
who have it to those who need it. This will be accomplished
through the use of the master index and the establishment of
standards and formats for recording and disseminating data.

Another agency is developing a central index to improve
access to, and exchange of, environmental data. NOAA
operates a number of environmental data and information
centers under one of its major components, the Environmen-
tal Data Service (EDS). EDS' major data centers are the
National Climatic Center, the National Oceanographic Data

Center, the National Geophysical and Solar-Terrestrial Data Center, and the Environmental Science Information Center.

According to an EDS study, each data center has developed separately from other centers and, until recently, these centers have appeared to be entirely separate entities with only a common administrative thread linking them to an environmental service organization. To provide users and data contributors access to data on a one-stop basis, EDS is developing a central index which is to identify the various types of environmental data available in the EDS centers and NOAA-type data available in other Federal agencies. The Environmental Data Index (ENDEX) is to contain detailed indexes to the data files of the EDS centers and should allow users timely access to data.

PROBLEMS IN EXCHANGING DATA

We asked managers to identify problems their systems experienced in acquiring data or converting and integrating data received from other systems, both intraagency and interagency data. About 43 percent of the managers responded to this question and cited the following problems.

Problems	Frequency of responses
Noncompatibility of data	67
No timely response	34
Data unreliable	31
Noncompatibility of equipment	29
Costs prohibitive	18
Other	29

Noncompatibility of data

Discussions with various systems' managers revealed that the manner in which environmental data is coded in the various systems hampers or prevents its exchange. Data is collected for a single purpose in one type of coding arrangement, and it is difficult to use that data when it is transferred to other systems having different codes. To overcome this problem, some managers have indicated that they must recode the data received from other sources according to their own individual system's configuration before the data can be used.

7

This problem has been noted in many studies of information and data programs. For example, the SEQUIP Committee noted that:

"It is obvious that standards will have to be established so that systems and data originating in one agency can be linked to similar systems in other agencies and eventually into a network of systems."

Through the use of standards, the Committee noted that:

"* * * information programs will be able to use compatible information and data files created by other organizations, thus reducing the overall data collection needs and concomitantly the total internal resources required."

In our May 16, 1974, report to the Congress entitled "Emphasis Needed on Government's Efforts to Standardize Data Elements and Codes for Computer Systems," (B-115369) we noted that:

"Communication barriers resulting from different agencies codes make it difficult and often impossible to consolidate data from different information systems. For example, the Civil Service Commission (CSC) found that agencies could not economically and accurately comply with its seemingly simple request for the total number of Government employees of each sex.

"Agencies were asked to code males '1' and females '2' and to provide the data on magnetic tape. That data was readily available, but the agencies defined and coded the data differently. For example, agency A combined sex data with education data, agency B combined sex data with marital status data. Agency C simply recorded Mr., Miss, Mrs., and Dr. It arrived at sex statistics by assuming that all or most doctors were male.

"CSC could not obtain the data it needed without going through a costly manual operation to convert the nonstandard data. As a result, CSC initiated a program to standardize data in the Federal personnel systems."

Also in our May 16, 1974, report we assessed the progress being made under the Federal program to standardize data elements and codes used in computer operations. We found that, since the program began in 1965, Federal efforts have been slow and not very successful due to the program's low priority and limited resources.

No timely response

According to the systems' managers, unavoidable delays occur in preparing the raw data and converting it into some type of usable format. Once a system has the raw data, it may take time to assemble, collate, and summarize it for distribution.

The SEQUIP study noted another problem causing delays in getting access to data. The study stated:

"* * * that information and data programs whose primary task is the support of their parent agencies are often reluctant to advertise their services and products outside their own agencies * * * because they fear being swamped with requests for these services without having resources to comply with such requests on a regular basis."

An official at the National Referral Center of the Library of Congress noted one problem limiting access to existing data. He explained that the priority placed on filling data requests from external sources is lower than other activities, because the primary missions of most Federal agencies are other than supplying information. As a result, agencies tend to cut funds from information activities first.

Data unreliable

Some managers indicated that the accuracy and reliability of data generated by other systems constitutes a problem. One system manager commented that no uniform standardization or degree of accuracy for data exists—nationally or internationally. He made the further observation that once data is received from other systems, one must have an indication as to how often the equipment collecting the data is checked for accuracy.

One manager noted that his system received data from 75,000 point sources. To check the accuracy of these sources would be a considerable task. Another manager indicated that the data can be unreliable because funds are

not always available for the proper maintenance and calibration of equipment.

Noncompatibility of equipment

In discussions with various managers, we found that the problem of noncompatibility of equipment included both hardware and software.

Hardware

There is a variety of manufacturers of computers and associated equipment. Although the concepts used to store data and programs on this equipment are similar, the actual methods used are sufficiently different to preclude direct interchange. For example, some managers said that the format in which data is stored on magnetic tape can vary and prevents direct exchange of tapes between systems. Another manager said that his IBM System 7 computer has both disc and magnetic tape drives which are incompatible with other computers and therefore prevents direct exchange of tapes or discs.

Software

Because there are many different computer program languages, there are differences both in the manner in which programs are written and in the manner in which data is processed and stored. Consequently, data recorded by a computer using one language may or may not be directly usable to a program written in another language.[1] For example, a manager using a UNIVAC 1108 indicated that his system uses certain levels of the FORTRAN language and that his system could use data directly from another system only if it used the same FORTRAN version.

The Federal Government has been attempting to increase the compatibility and interchangeability among computer systems. The Secretary of Commerce is authorized by law to make

---

[1]Data could be recorded in two systems using the same data codes, that is, "A" might represent "male" in both systems. However, the method of recording the letter "A" might be different. Data is recorded in most computer systems using a series of on or off magnetic pulses and the arrangement of these pulses can differ. For example, "A" might be stored as magnetic pulses 0001 in one system and as magnetic pulses 1001 in another system.

10

recommendations to the President to establish uniform standards, referred to as Federal Information Processing Standards. An objective of the standards is to facilitate the interchange of data and programs.

In this regard, in March 1968, the President approved the Secretary's recommendation to establish a standard Code for Information Interchange. This standard requires that Federal computers be able to use a specific code and character set in storing data to improve their ability to exchange data.

Although the problem of noncompatibility of equipment is complex, techniques are available to solve these problems and facilitate program and data exchange.

These techniques involve the use of software conversion programs which will convert data from one format to another. The computer industry has developed software programs and small computers whose sole function is to perform this conversion process.

Although conversion of data is possible, these conversions could be very expensive, depending on the degree of incompatability, the volume of data, and how soon the data is needed.

Costs prohibitive

System managers have indicated that cost is a problem associated with the noncompatibility of equipment and data. Some systems do not have the funds or must incur added expense to convert the data received from other systems into a format that can be readily used. For example, one system prefers to receive its data on magnetic tape in a prescribed format but, since 60 percent of the incoming data is in hard-copy form, it must incur the expense of transcribing the data onto magnetic tape.

DIRECTORY OF ENVIRONMENTAL DATA SYSTEMS

The 1971 SEQUIP Committee noted the difficulty in acquiring environmental information. According to the Committee, even managers of information and data centers in Federal agencies often did not know that similar systems and data banks exist in other agencies and sometimes in their own departments. The SEQUIP Committee Chairman told us that system managers were surprised to learn that data they were

collecting was also being collected by other organizations. The Committee further asked the rhetorical question "How much more difficult is it for an outsider--the Congressman, the municipal agency manager, the academic scientist--to know just where to obtain needed information."

As a result, the SEQUIP Committee recommended that the National Referral Center compile an environmental information resources directory. However, the Center has not published a directory, primarily because of a lack of funds.

During September 1972, EPA sponsored a National Environmental Information Symposium in Cincinnati, Ohio. As noted in its summary report:

> "The most common concern expressed by Symposium participants, regardless of attitudinal or professional orientation, was the need for improved awareness of, and access to, environmental information."

It appears that the lack of awareness of what data has been collected and by whom is still a problem. Managers of 151 systems responding to our questionnaire expressed the need for an overall directory of environmental data systems.

In addition, 56 other managers indicated that a directory was not needed and some of these managers expressed the belief that (1) their organizations could not use a directory, (2) they were the only ones collecting special types of data, or (3) they were aware of pertinent data systems.

Those managers who expressed a need for a directory indicated that a directory of environmental data systems could

--eliminate duplication in the collection of data,

--reduce search time and the number of referrals, and

--identify existing information sources and what data is available.

We asked system managers to indicate the types of information they would like in a directory of environmental data systems. Following is a summary of the items most frequently mentioned.

12

1. Name and description of system.

2. Method to gain access to the data (person and agency to contact).

3. Scope and kinds of data available.

4. Reliability and quality of data.

5. Methods of storing and manipulating the data, such as computers and computer languages.

Existing directories

Some agencies had prepared directories of their data systems. EPA has compiled and published a directory of its environmental information systems. Also, certain managers of data systems at the Food and Drug Administration (FDA) noted that FDA had published a directory entitled "Computer Systems in FDA."

There are organizations which have compiled directories, listing sources of environmental information beyond the scope of a single agency or organization. However, no Federal agency is maintaining a directory that could be considered a complete directory of all environmental data systems.

For example, the Council on Environmental Quality issued a Federal Environmental Monitoring Directory in May 1973. This directory lists primarily Federal data systems that are involved in environmental monitoring--the systematic and continuing observation of environmental parameters--and that are collecting data nationwide. Other systems, such as systems monitoring environmental research information, would not be listed in the Council's directory but are important sources of environmental information.

According to a Council official, the Council has tentative plans to update this manual but not until, at least, 1976. The directory has not been computerized. Council officials noted that it would be difficult to update the directory, because rapid changes occurring in the environmental area will result in changes in the sources of environmental data.

Plans for developing directories

The Department of the Interior is planning to develop directories that will list references to certain

environmental data sources. However, these data directories will list data sources related to a specific functional area but will not list all environmental data sources.

The Department of the Interior is developing a Resource and Land Information (RALI) program to improve Federal capabilities in land use and natural resource planning and to provide a continuing referral service to the information resources of the United States pertaining to land use and resources development. The RALI program will develop and maintain catalogues, directories, and indexes. Much of the data sources compiled in those catalogues and directories is to be environmental-type data.

As noted earlier, the Department of the Interior is also developing NAWDEX to facilitate, by use of a central index and referral system, the transfer of water data from those who have it to those who need it. A directory is to be published from NAWDEX's central index.

National Referral Center

The National Referral Center maintains an extensive index of information resources. The Center publishes directories of data sources which include professional societies, university research bureaus, Federal and State agencies, industrial laboratories, and individual experts as well as the more traditional sources of information, such as technical libraries, information and documents centers, and abstracting and indexing services. Although present catalogues contain numerous references to environmental data sources, an environmental data catalogue has not been published. A National Referral Center official told us that the Center had considered reorganizing its present files to enable it to issue a directory. The official explained that the Center had no plans to do so, primarily due to limited funds. However, the official said that the Center planned to issue a directory of water data sources and was considering directories in other environmental areas, such as air.

According to a National Referral Center official, sources of information listed in the Center's directories do not generally identify specific data files or systems and therefore users cannot determine the extensiveness of the data maintained by the organization. However, the Center is atttempting to identify the names of specific data systems or files that are maintained by the various information sources and is to include this information in future revisions of its directories.

CHAPTER 3

LEGISLATION WHICH WOULD ESTABLISH A

NETWORK OF ENVIRONMENTAL DATA SYSTEMS

A number of bills have been introduced in the Congress to establish a central organization to coordinate the collection and exchange of environmental data. One bill--House bill 56--was passed by both Houses of the 92d Congress but was vetoed by the President. He said it would establish a central computer system which would be unnecessarily costly because it would lead to duplication of information or produce results unrelated to real needs. Another bill introduced in the 93d Congress--House bill 36--would provide for a network of new and existing data processing facilities which through a system of interconnections would be in communication with a central facility. As of September 1974, the Congress had not acted on House bill 36. The Subcommittee Chairman asked us to look into the costs and problems of establishing an environmental data network.

LEGISLATION

House bill 56

In 1972 House bill 56 was passed by both Houses of Congress to amend the National Environmental Policy Act of 1969 by establishing a National Environmental Data System and creating State and regional environmental centers. Section 102(b) of the bill stated:

> "The purpose of the Data System is to serve as the central national coordinating facility for the selection, storage, analysis, retrieval, and dissemination of information, knowledge, and data relating to the environment so as to provide information needed to support environmental decisions in a timely manner and in a usable form. Such information, knowledge, and data as shall be deemed appropriate and useful for the achievement of the purpose of the system shall be made available by all Federal agencies and shall be collected and received, where available, from all Federal agencies, private institutions, universities, and colleges, State and local governments, individuals, and any other source of reliable information, knowledge, and data."

House Report 92-203 issued by the Committee on Merchant
Marine and Fisheries on May 13, 1971, and accompanied House
bill 56 noted that:

> "The term 'Data System' shall be construed to
> include an appropriate network of new and existing
> information processing or computer facilities
> throughout the United States. The Data System
> would be developed and established and consist
> of a central facility capable of interconnecting
> and communicating with other systems and equip-
> ment * * *."

In the President's October 21, 1972, Memorandum of
Disapproval of House bill 56 he noted that the creation of
the data system and the environmental centers would lead "to
duplication of information or would produce results
unrelated to real needs and wasteful of talent, resources,
and the taxpayers' money." The President further stated:

> "I believe there are serious drawbacks to such a
> data system which would outweigh potential
> benefits. The collection of data and statistics
> on the supposition that some day they may be useful
> is in itself a highly dubious exercise. Data, taken
> out of the context of the questions they were
> specifically designed to answer, can even contri-
> bute to confusion or be misleading.

> "With this in mind, I believe the centralized col-
> lection of environmental data should be related to
> specific policies and programs. H.R. 56 fails to
> provide such a relationship * * *."

House bill 36

On January 3, 1973, the Chairman of the Subcommittee on
Fisheries and Wildlife Conservation and the Environment,
House Committee on Merchant Marine and Fisheries, introduced
House bill 36 which was similar to House bill 56 in that it
would establish a National Environmental Data System to
serve as a central facility for the coordination of existing
and proposed environmental data systems and programs.

House bill 36 specifically states that the data system
would be a

"* * * network of new and existing information
processing or computer facilities both private
and public * * * which, through a system of inter-
connections, are in communication with a central
facility * * *."

The bill would establish a director of the data system
who would be qualified to analyze and interpret environ-
mental data of all kinds. The director would also be
responsible for developing predictive ecological models.
Environmental data and analysis of data would be available
to the Congress and Federal departments and agencies as well
as interstate agencies, and State and political subdivisions
thereof.

The Chairman of the Committee on Merchant Marine and
Fisheries asked us to comment on House bill 36. In our
comments we expressed our belief that the bill would fulfill
the need for a national environmental data system which
would coordinate the independent data systems of the many
Federal agencies conducting environmental research and
development programs. Our comments were based upon a review
of the research and development (R&D) programs related to
the prevention and control of water pollution.

In that review we noted that a number of Federal
agencies were collecting and disseminating water pollution
R&D information and that their data systems were not
coordinated and were not as useful as they might have been
to those interested in the results of water pollution R&D
efforts. We noted that several major Federal systems were
disseminating R&D information related to water pollution;
however, none were complete or comprehensive in coverage and
there was a lack of effort by those groups responsible for
gathering information to identify the users of research data
and their needs. In our report to the Congress entitled
"Research and Demonstration Programs to Achieve Water
Quality Goals: What the Federal Government Needs to Do"
(B-166506, Jan. 16, 1974), we recommended that the Office of
Management and Budget (OMB) designate a Federal agency as a
focal point to coordinate and promote the dissemination of
water pollution research results. An OMB official on
September 20, 1974, said that OMB was considering the
recommendation but had not taken any action on it.

COSTS AND PROBLEMS OF ESTABLISHING
AN ENVIRONMENTAL DATA NETWORK

In January 1974 the Chairman asked that we develop (1) cost information for an existing computer network for the purpose of relating it to the establishment of a computer network for coordination and exchange of environmental data and (2) information on the problems of establishing such a network.

The National Institutes of Health (NIH), Department of Health, Education, and Welfare, operates the Computer Center which has four large central processing units in communication with 71 remote terminal computers and 750 remote conversational typewriter terminals.

The Director of the Computer Center told us that the costs of the NIH network or any other network could not be readily related to the costs of establishing a computer network for environmental data. He said that, to arrive at costs for a new network system, one must first determine what the specific purposes are and what is to be accomplished. On the basis of the various needs and requirements, a network design could be determined which would dictate the types of hardware and software needed and thereby reliable cost estimates could be developed. For example, the cost of hardware is linked to the storage and job-processing capacity which is related to the type of work. Also it would cost more to hold data "on line" for immediate access by a central computer facility than if the data were stored, for instance, on magnetic tape in a tape library; the greater the volume of on-line data, the greater the cost of the network.

As we have found, there are other less costly methods of transferring data between data systems than by direct interconnection of computers. For example, data stored on magnetic tape could be sent to another system through the mails. One situation where the interconnection of computers may be desirable would be when data is valuable only if it is received within a short time frame.

We discussed with the Director of the RALI program in USGS the problems of developing a data network. A data network was being considered for the RALI program to support land use and resources planning at the Federal, State, and local levels. The concept for the network has a number of similarities with House bill 36.

The RALI program concept was submitted to USGS in a October 1972 report by its contractor. Under this concept, the RALI program would have made maximum use of existing information sources by providing (1) references to data maintained by other organizations and (2) a computerized information base for modeling and analysis tasks consisting of information summarized from data gathered and maintained by other organizations.

As part of the RALI concept, a National Information Facility would have been established "to serve the analysis and data needs of the Federal establishment primarily on a national level." The facility would have supplied summarized data to Federal agencies, legislative and judicial units, and other national-level planning groups to allow these groups to make broad predictions of the probable impact of national policy decisions on land use and national resources.

According to the report, the RALI program would have also established a number of regional, State, and local information centers to provide referencing to and researching of available data needed by users.

Most of the data available to users would not have been an integral part of the program's data files but would have been that data collected and maintained by the agencies, governments, or organizations whose charter, objectives, or capabilities place these acquisition and storage functions most clearly within their domain.

One of the major benefits of the RALI program could have been the elimination of duplication in collection of data in that only one data gathering would be performed and the data would have then been available to all users.

The report suggested that, before full implementation, the RALI program be developed in four phases--the conceptual phase, the definitive phase, the development phase, and the initial implementation phase. The definitive phase would provide the major source of information needed to develop RALI program design alternatives. During this phase, which would have taken about 1-1/2 years, users needs and data sources which could have been incorporated into the RALI program would have been identified as well as any modeling and other analysis tools in operation or under development.

The Director of the RALI program told us, however, that USGS could not implement the report's suggestions in the

definitive phase. He explained that the report's approach was data oriented rather than problem oriented.

He said that most available data was useful primarily for the specific and limited purposes for which it was collected and might have only marginal utility beyond its original use. An example is digital slope data collected by the military for cross-country maneuvering purposes. Terrain is divided, by percentage of slope, into such categories as suitable for wheeled vehicles, suitable for tracked vehicles, and impassible for any vehicle. Other groups, such as foresters, land use planners, construction engineers, railroads, and highway builders, may collect slope data for the same geographic area, but their slope categories will differ and will not be interchangeable with the military or with each other.

Slope is an example of derived data. All groups begin with measurement of elevation--basic data--but each group then computes and aggregates it into different categories based on percentage of slope. The basic data has many uses but only after it has been processed into another form. The derived data is directly usable but only to the group that compiled it. Because most data is collected for some specific purpose, the collections tend to be derivative in nature and have limited transferability. He said the problem was not so much one of sharing data but was one of determining what data can be usefully shared.

He pointed out that collecting and maintaining data is very costly and that data should not be introduced into an information system without regard to its overall use. He added that an information network should not be designed around miscellaneous, unrelated, and limited-use data as collected by various data users but should be designed around the problems to be solved. Once the problems are identified, the various data needed to solve the problems can be determined.

The Director of the RALI program noted that no decision had been made regarding the use of the technique of interconnecting or networking computers because it was not appropriate this early in the program's developmental stage. He said that the interconnection of computers would be decided at a later time on an individual case basis when the need for networking is justified. He noted that in many cases other techniques of sharing data, such as exchange of magnetic tapes, may be more economical.

CHAPTER 4

CONCLUSIONS AND MATTERS FOR

CONSIDERATION BY THE SUBCOMMITTEE

CONCLUSIONS

Many Federal agencies collect and store environmental data, in many instances in the same functional area, such as air and water. These data systems appear to have developed independently of each other to meet the specific missions of the various agencies. However, although there is no central focal point for coordinating the establishment of such systems or for the exchange or collection of data, many systems share information with other Federal agencies. A number of problems hindered the exchange of data, including a lack of uniformity in the way data is coded and a lack of awareness of what data has been collected and by whom. One possible solution to the problem of a lack of awareness could be the creation of a directory of environmental data systems. Many systems managers agreed that such a directory was needed.

Legislation introduced in the Congress would establish a central focal point for coordinating the collection and exchange of environmental data. The legislation would also provide for a computer network of new and existing data processing facilities which through a system of interconnection would be in communication with a central facility. There are, however, other less costly methods of sharing data than through the interconnection of computers. Our review has shown that some data systems are sharing data with other Federal agencies through the exchange of some computer media, such as magnetic tape.

We believe that, if a network of environmental data systems is to be established, it should be designed around the environment problems that need solving. First, it would be necessary to define the environmental problems that are to be addressed and to determine the analysis tools, such as predictive models which would assist in their solution. Defining the problems would involve considering their ecological, political, and social aspects. Once the problems and tools are identified, the data to be held in such a network can be identified and the best method of storing and exchanging such data can be determined.

MATTERS FOR CONSIDERATION BY THE SUBCOMMITTEE

We suggest that the Subcommittee, when considering House bill 36, specify that, before a nationwide network of interconnected environmental data systems is established, the environmental problems to be solved must be defined and the analysis tools needed to assist in solving the problems be determined. Once the problems and analysis tools are identified, the data needs for the network and the best methods for storing and exchanging such data can be determined.

We also suggest that the central organization, which would be established if the bill were enacted, be made responsible for establishing and maintaining an environmental data directory to increase awareness of what data is available.

U.S. House of Representatives
Committee on
Merchant Marine and Fisheries
Room 1334, Longworth House Office Building
Washington, D.C. 20515

March 8, 1973

Honorable Elmer B. Staats
Comptroller General of the U.S.
General Accounting Office
Washington, D. C. 20548

Dear Mr. Staats:

Environmental data and information are being collected by a number of diverse Federal organizations. Virtually every department and a dozen offices and councils within the Executive Branch collect and store vast quantities of data. In addition, at least 24 independent agencies involved in environmental research or management collect similar amounts of data. These data are a national resource upon which we must depend for information in guiding our efforts to protect the environment. This information represents an investment of several million dollars and is an irreplaceable asset.

I am concerned that we are not making maximum use of this valuable resource. There is no systematic directory of the environmental information collected and stored by the Federal Government, nor does there appear to be any degree of uniformity in the way it is stored and disseminated. Perhaps more importantly, I suspect that there is little coordination among the agencies with regard to the types of environmental data collected and the exchange of information among governmental entities.

Thus I am requesting GAO to initiate an audit of the Federal agencies charged with the collection of environmental data, or which incidental to their missions collect environmental information, for the purpose of evaluating the following points:

* Kinds of environmental data collected and scope
  of the data collection programs.

* Method of storing information.

* Availability of information to private and public
  sectors, including other government agencies.

* Methods of coordinating data collection and
  transfer among other government agencies.

Please compile a report based on the findings concerning
the points mentioned above and include any legislative pro-
posals you deem appropriate as a result of your audit.   I
suggest that your representative confer with the staff of my
Subcommittee on Fisheries and Wildlife Conservation and the
Environment and the staff of the Environmental Policy Divi-
sion, Congressional Research Service, Library of Congress,
for further details regarding the study.   If feasible, I
would like to have the report available by November 1, 1973.

Thank you for your help with this matter.

Sincerely,

John D. Dingell, Chairman
Subcommittee on Fisheries
and Wildlife Conservation
and the Environment

SUMMARY OF RESPONSES TO GAO QUESTIONNAIRE

ON FEDERAL ENVIRONMENTAL DATA SYSTEMS

We distributed a questionnaire to those Federal agencies and departments that we identified as being involved in the collection, storage, and/or dissemination of environmental data. We did rely, to a certain extent, on each agency to determine which systems or programs had environmental data.

We received 320 completed questionnaires from 18 departments and agencies. The following section summarizes the responses obtained for each of the four points raised by the Chairman.

KINDS OF ENVIRONMENTAL DATA COLLECTED
AND SCOPE OF THE DATA COLLECTION PROGRAMS

Many of the systems are collecting data in the same functional areas as shown by the following responses.

Functional area	Number of responses
Water	175
Air	120
Land use	105
Pesticides	62
Noise	27
Radiation	49
Solid waste	45
Other	81

Some responses indicated that many systems were collecting data in more than one functional area. For example, one system was collecting data on pesticides in water.

In the "other" category, respondents indicated a variety of specialized areas, such as plant and animal data.

We asked each respondent to indicate the primary purpose of his system's data collection efforts. Some respondents indicated more than one primary purpose for their system, and their responses are summarized in a separate column.

	Number of responses	
Purpose of data collection	Primary purpose	Two or more purposes
Legal, legislative, or regulatory	33	35
Surveillance and monitoring	49	40
Research and development	53	35
Management and planning	85	52
Socioeconomic	4	14
Other	25	5

Many respondents indicating "other" purposes indicated a special purpose which they believed did not fit into the above categories. A number of these respondents noted that they were depositories for data and stored information for use of other programs and organizations.

We asked each respondent to indicate whether his system was primarily data oriented or information oriented. Data-oriented systems are concerned with obtaining facts, principally in digital forms; for example, systems involved in surveillance and monitoring activities. Information-oriented systems are concerned with data in narrative form, such as reports, journals, and specialized or technical literature. Of 283 systems managers responding to this question, 158, or 56 percent, said their systems were primarily data oriented and 125 were evenly distributed between information oriented and "other." The remainder of the respondents indicated two or more categories or did not answer the question. "Other" included chiefly data collection techniques, such as sampling, questionnaires, canvassing, or onsite visits and evaluations.

We asked each systems manager to describe the range and extent of his data collection activities. We categorized the scope by geographic area for 266 systems. The responses for the other systems could not be classified by geographic area.

Scope of information or data	Number of systems	Number of systems indicating data in each functional area						
		Air	Water	Land use	Pesticides	Noise	Radiation	Solid waste
International (note a)	58	34	41	16	13	9	11	11
National (note b)	128	30	58	51	23	6	11	14
Regional (note c)	80	28	46	22	7	3	11	5
Total	266	92	145	89	43	18	33	30

aSystems collecting and storing data worldwide or from selected areas outside and in some cases inside the United States and its territories.

bSystems collecting and storing data throughout the United States.

cSystems collecting and storing data from selected areas of the United States, such as the Southeastern States or a particular river basin.

STORAGE OF INFORMATION

Environmental information is stored totally, or in part, in computer format by 229 systems. Some of these systems also maintain part of their data in written reports and publications or manual files. Only 90 systems managers indicated that they did not store any of their data in computer media. Other methods of storage the respondents noted were microfilm, photography, maps, and charts.

We asked the managers using a computer media for storage of all or part of their data to indicate the media. The following schedule shows the number of responses for each media.

Computer media	Number of responses
Punched cards	104
Magnetic tape	187
Magnetic disc	115
Magnetic drum	7
Punched tape	20

The responses to our questionnaire revealed that these systems used a variety of computer models from many different manufacturers to process environmental data. The following schedule shows the makes and number of different models used.

Make	Number of models	Number of systems used
IBM	24	160
Control Data Corp.	9	42
DATA 100	2	4
UNIVAC	4	18
Digital Equipment Corp.	2	8
Honeywell	7	11
Burroughs	2	7
Modular Computer Systems	1	1
Wang	1	1
Xerox	4	6
Data General	1	4
Varian	1	2
Total	58	264

The data systems used a variety of computer languages.

Language	Number of responses
FORTRAN	140
COBOL	89
RPG	11
BASIC	14
PL/I	46
Assembly (various)	40
MARK IV	9
IRS	10
Misc.	34

Many systems use two or more languages.

AVAILABILITY OF DATA

Respondents indicated that they provide their data to organizations in other agencies and departments in the Federal Government. The following schedule shows the number of respondents who indicated their systems provide data to organizations outside their own departments or agencies.

Agency	Number of respondents	Number providing data to other departments and agencies
Agriculture	33	17
Commerce	39	31
EPA	87	22
Interior	61	37
Atomic Energy Commission	23	14
Corps of Engineers	10	8
Health, Education, and Welfare	18	7
National Science Foundation	10	5
Tennessee Valley Authority	11	9
Others	28	17
Total	320	167

The system managers also indicated that they provide data to organizations outside the Federal Government which also constitutes a large portion of users of environmental information. The following schedule indicates the major non-Federal recipients of environmental information and the number of systems which provide information.

Recipients	Number of systems providing information
State	163
Academic community	137
Private sector	139
Business and industry	49
Local governing bodies	36
Foreign countries	13

Systems managers were requested to indicate how intraagency and interagency Federal users and any other users outside the Federal Government get data from their systems. The following schedule shows the number of responses in each category.

Computer terminal	86
Telephone	151
Mailing list	129
Other	165

Respondents indicating "other" noted (1) various methods of receiving requests, such as direct personal contacts and written correspondence, or (2) methods of transferring data or information, such as general and annual reports, computer printouts, publications, journals, and magnetic tapes.

Most respondents indicated that their systems receive 50 or fewer requests a month for information as shown by the following schedule.

Number of request for information	Number of responses
0 to 50	209
51 to 100	35
101 to 500	25
501 or over	27

We asked each systems manager to indicate a normal response time for requests for information from his system.

Responses received from 250 managers are shown below.

Immediate or 1 to 2 days	3 days to less than 2 weeks	2 weeks to less than 1 month	1 month to less than 2 months	2 months to less than 6 months	6 months and over
85	92	41	19	6	7

Approximately 71 percent of the respondents indicated that their normal response time was less than 2 weeks.

Most respondents (230) indicated that they did not charge users for the services they provide. The primary reasons given for not charging users were: (1) services were for intraagency use only, (2) information was given free as a public service, (3) information was given free except when unusual processing was necessary.

METHODS OF COORDINATING DATA COLLECTION AND TRANSFER AMONG GOVERNMENT AGENCIES

To determine the various methods of coordination in data collection and transfer among the agencies, we asked each respondent to indicate (1) if his system participated in a network which has as its purpose either the exchange of data or the coordination of data collection activities and (2) if his system received environmental data from other systems. The schedule below shows those agencies indicating the largest number of networks for exchange or coordination of data.

Agency	Number of respondents	Number indicating involvement with a network
Agriculture	33	11
Commerce	36	20
Atomic Energy Commission	20	10
Interior	58	13
EPA	86	21
All others	70	23
Total	303	98

We asked each systems manager who indicated participation in a network to describe the network. Most of the explanations were very brief and did not explain how data was exchanged or coordinated.

Of those responding, 141 indicated their systems received environmental data from other systems. The schedule below indicates those agencies having the largest number of systems which receive environmental data from other systems.

Department or agency	Number of systems receiving data from		
	Intra- agencies	Interagencies	Both intraagencies and interagencies
Agriculture	4	5	1
Commerce	10	11	11
EPA	16	8	3
Interior	7	9	4
Atomic Energy Commission	0	5	9
All others	2	22	6
Total	39	60	34

Eight respondents did not indicate the source of the data.